

Governança preditiva e viabilidade de um port community system (pcs) no porto de santos: uma abordagem de ciência de dados para a tomada de decisão compartilhada



Predictive governance and feasibility of a port community system (pcs) in the port of santos: a data science approach for shared decision-making.

Robert Richard das Neves Correia Santos 

Fatec Santos
robert.richard.scripts@gmail.com

Juliana Brondino 

Fatec Santos
brondinoo@gmail.com

Sandra de Oliveira Soares Cardoso 

Fatec Santos
s.oliveira61@gmail.com

José Augusto Theodosio Pazetti 

Fatec Santos
josepazetti@gmail.com

Revista Processando o Saber

eISSN 2179-5150 · Vol 18, n. 01, 2026
Multidisciplinar · DOI · Revisão por pares

Faculdade de Tecnologia Praia Grande – FATEC

Períodicidade: Anual
revista@fatecpg.edu.br

Recebido: Jan 2026

Aceito: Mar 2026

Publicado: Jun 2026

URL: <https://www.fatecpg.edu.br/revista/index.php/ps/article/view/459>

DOI: <https://doi.org/10.5281/zenodo.20074664>



RESUMO

Este artigo apresenta um estudo de viabilidade para a implementação de um Port Community System (PCS) no Porto de Santos, focando em como a análise preditiva pode mitigar conflitos de governança e barreiras ao compartilhamento de dados. O setor portuário brasileiro apresenta elevada fragmentação informacional, impactando a eficiência do comércio exterior nacional. A metodologia fundamenta-se no ciclo de vida da Ciência de Dados, utilizando uma integração inédita de bases de dados da ANTAQ, Porto Sem Papel (PSP) e Estatísticas de Anuências. Técnicas de pré-processamento como KNN Imputer, One-hot Encoding e Standard Scaler foram aplicadas, seguidas por um modelo preditivo XGBoost otimizado via framework Optuna para analisar gargalos de Lead Time. Os resultados, validados por valores SHAP e análise de clusters (K-Means), demonstram que a transparência gerada por modelos preditivos atua como um mediador de confiança, reduzindo a entropia informacional e facilitando a governança compartilhada entre entes públicos e privados. O estudo propõe um framework de integração que utiliza a análise de dados para fundamentar a tomada de decisão compartilhada, visando elevar a competitividade do complexo santista.

PALAVRAS-CHAVE: Inteligência Artificial; Educação; Tecnologia; Aprendizagem; Ética.

ABSTRACT

This article presents a feasibility study for the implementation of a Port Community System (PCS) at the Port of Santos, focusing on how predictive analytics can mitigate governance conflicts and data-sharing barriers. The Brazilian port sector presents high informational fragmentation, impacting the efficiency of national foreign trade. The methodology is grounded in the Data Science lifecycle, integrating datasets from ANTAQ, Porto Sem Papel (PSP), and Clearance Statistics. Preprocessing techniques such as KNN Imputer, One-hot Encoding, and Standard Scaler were applied, followed by an XGBoost predictive model optimized via the Optuna framework to analyze Lead Time bottlenecks. The results, validated by SHAP values and cluster analysis (K-Means), demonstrate that predictive transparency acts as a trust mediator, reducing informational entropy and facilitating shared governance between public and private stakeholders. The study proposes an integration framework that utilizes data analysis to support shared decision-making, aiming to enhance the competitiveness of the Santos port complex.

KEY-WORDS: Artificial Intelligence; Education; Technology; Learning; Ethics.

INTRODUÇÃO

Os *Port Community Systems* (PCS) são definidos como plataformas eletrônicas que permitem o intercâmbio inteligente e seguro de informações entre stakeholders públicos e privados, sendo fundamentais para a competitividade (SIMÕES et al., 2024).

Trabalhos anteriores desta série identificaram os benefícios dos *Port Community Systems* (PCS) e mapearam nove categorias de boas práticas essenciais para sua implementação.

Embora o Porto de Santos figure entre as organizações mais eficientes do país (MINISTÉRIO DE PORTOS E AEROPORTOS, 2025), a implementação de um PCS enfrenta barreiras que transcendem a tecnologia, recaindo sobre a governança de dados. Questões como a propriedade e o controle das informações, bem como a arbitragem de impasses entre empresas concorrentes, representam os principais entraves para a sustentabilidade do sistema. No modelo de governança portuária *Landlord Port*, a autonomia e uma estrutura de decisão compartilhada são fundamentais para que o compartilhamento de dados de negócio seja visto como um valor estratégico e não como uma exposição de vulnerabilidades.

Este artigo busca preencher a lacuna entre a teoria e a prática, fundamentando-se nos achados de Simões et al. (2024), que identificaram os benefícios operacionais dos PCS, e de Simões et al. (2025), que mapearam a "Governança e Colaboração Multissetorial" como a boa prática de maior peso para o sucesso dessas plataformas. A fragmentação informacional e a burocracia documental, identificadas como barreiras técnicas e organizacionais, são aqui tratadas não apenas como problemas de TI, mas como falhas de governança que exigem visibilidade analítica para serem superadas.

O objetivo central é demonstrar a viabilidade de um PCS no Porto de Santos por meio de uma Governança Preditiva. Utiliza-se o algoritmo XGBoost para prever o Lead Time e, por meio da transparência analítica (SHAP Values), busca-se fornecer subsídios para a tomada de decisão compartilhada. A proposta é transformar dados brutos em argumentos para a colaboração, mitigando a resistência cultural ao compartilhamento de informações entre os *stakeholders*.

1. FUNDAMENTAÇÃO TEÓRICA

A fundamentação teórica deste estudo explora a convergência entre a governança e o ciclo de vida da Ciência de Dados, estabelecendo as bases conceituais para que a integração de *stakeholders* e a análise preditiva atuem como pilares de sustentabilidade e eficiência em um ecossistema PCS.

1.1 A RELEVÂNCIA DA GOVERNANÇA E COLABORAÇÃO

A "Governança e Colaboração Multissetorial" é o fator de maior peso para a sustentabilidade de um PCS. Conforme apontado por Simões et al. (2025), esta categoria de boa prática é a mais citada na literatura, exigindo uma estrutura de decisão compartilhada e a definição clara de papéis entre os atores. No complexo santista, que opera sob o modelo de governança portuária *Landlord Port*, a multiplicidade de agentes envolvidos demanda soluções que mitiguem a fragmentação de dados e a assimetria de informação.

Nesse modelo, embora a autoridade portuária exerça o papel regulador, a propriedade e o controle dos dados de negócio permanecem com os entes privados, gerando impasses sobre o compartilhamento de informações entre concorrentes. A implementação de um PCS eficaz permite o uso eficiente dos dados e a melhoria de processos, mas sua viabilidade depende da superação da resistência cultural ao compartilhamento.

A Governança Preditiva surge como uma hipótese para resolver esses conflitos: ao utilizar modelos de Ciência de Dados para prever gargalos como o Lead Time, o PCS deixa de ser apenas um repositório burocrático e passa a ser um mediador técnico. A transparência gerada por ferramentas analíticas atua como um facilitador para que as atividades ocorram de forma padronizada, transformando a colaboração em um valor central da comunidade portuária ao oferecer benefícios tangíveis de eficiência para todos os envolvidos.

1.2 CIÊNCIA DE DADOS COMO SUPORTE À DECISÃO

O Aprendizado de Máquina (Machine Learning), como um subcampo da inteligência artificial (IA), permite que sistemas aprendam autonomamente a partir de dados, sem necessidade de programação direta para cada tarefa específica. No contexto de um PCS, essa tecnologia não é apenas um recurso computacional, mas um instrumento de Governança

Digital, pois permite processar grandes volumes de dados heterogêneos para gerar indicadores imparciais de desempenho.

Para mitigar a fragmentação informacional no Porto de Santos, é necessário um pipeline robusto que inclua técnicas de padronização (Standard Scaler) e codificação (One-hot Encoding). Algoritmos de *boosting*, como o XGBoost, são eficazes para modelar relações complexas entre o volume de carga, a burocracia de anuentes e o Lead Time operacional. A otimização via Optuna automatiza a busca por *hiperparâmetros*, garantindo a precisão analítica necessária para que os resultados sejam aceitos pelos diversos *stakeholders*.

A integração dessas ferramentas de Ciência de Dados oferece o suporte técnico para a tomada de decisão compartilhada, uma vez que a explicabilidade do modelo (via Feature Importance e SHAP Values) permite identificar proativamente gargalos burocráticos. Assim, a ciência de dados atua como o motor de inferência que transforma a governança estática em uma Governança Preditiva, onde gestores podem atuar na redução dos tempos de anuência antes que estes impactem o fluxo final do complexo santista.

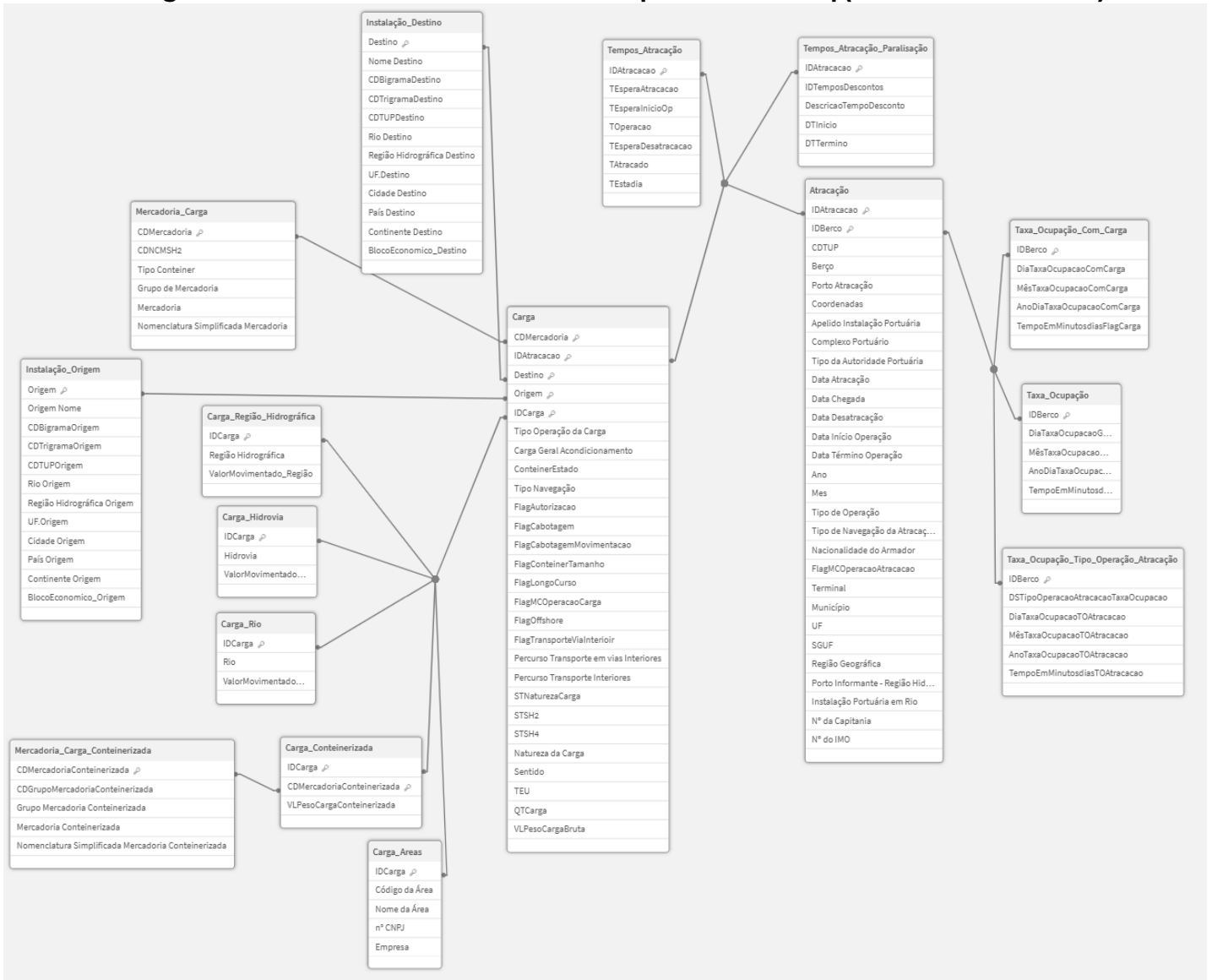
2. PROCEDIMENTOS METODOLÓGICOS

A metodologia deste estudo foi estruturada em um pipeline de ciência de dados executado inteiramente na linguagem de programação Python, utilizando o ambiente de desenvolvimento Google Colaboratory para garantir a reprodutibilidade. Este fluxo automático conduz os dados desde a integração de fontes heterogêneas até a geração de inteligência para a governança. As principais bibliotecas utilizadas foram pandas para a manipulação do *dataset*, Scikit-learn (PEDREGOSA et al., 2011) para pré-processamento e modelagem, Matplotlib (HUNTER, 2007) e Seaborn (WASKOM, 2021) para visualização, e Optuna para a otimização de *hiperparâmetros*.

2.1 CONSOLIDAÇÃO DO DATASET E ENGENHARIA DE DADOS

A Figura 1 apresenta o modelo de dados da coleta inicial dos dados, onde, posteriormente, foram utilizados registros de atracação, carga e tempos.

Figura 1 – Modelo de Dados - Estatístico Aquaviário – Antaq (atualizado abril/2025)



Fonte: Estatístico Aquaviário - Antaq (2025)

Para mitigar a fragmentação informacional citada por Simões et al. (2025), construiu-se um *dataset* multivariado inédito (2023-2025) integrando três fontes de dados distintas:

- Estatístico Aquaviário (ANTAQ): Registros de atracação, carga e tempos operacionais.
- Governança (Anuências): Variáveis de contexto burocrático, focando no tempo médio de análise de órgãos reguladores.
- Logística/Social (Porto Sem Papel): Dados de tripulação, passageiros e rastreabilidade via Documento Único Virtual (DUV).

A amostra final totalizou 17.364 registros únicos, com um *Lead Time* médio de 136,76 horas. Esta integração de bases é o primeiro passo para a soberania de dados do PCS, permitindo que a Autoridade Portuária tenha uma visão unificada do complexo.

2.2 PRÉ-PROCESSAMENTO DOS DADOS

Para garantir a integridade do modelo, aplicou-se um pipeline de pré-processamento no qual, após a limpeza inicial, a imputação de valores ausentes para variáveis como tripulação e peso foi realizada pelo método KNN Imputer ($n_neighbors=5$).

Na sequência, variáveis categóricas, a exemplo dos "Terminais", foram transformadas via One-Hot Encoding, técnica que converte categorias em representações numéricas fundamentais para o processamento pelos algoritmos (HASTIE; TIBSHIRANI; FRIEDMAN, 2009).

2.3 SEGMENTAÇÃO E REDUÇÃO DE DIMENSIONALIDADE

Aplicou-se o PCA (Principal Component Analysis) para reduzir a dimensionalidade, retendo a maior parte da variância original.

A análise de componentes principais, ou PCA, reduz o número de dimensões em grandes conjuntos de dados aos componentes principais que retêm a maior parte das informações originais. Ela faz isso transformando variáveis potencialmente correlacionadas em um conjunto menor de variáveis, chamadas componentes principais. (IBM, 2023, n.p.).

Em seguida, algoritmo K-Means agrupou as operações em 4 *clusters* distintos, permitindo identificar perfis de maturidade digital e eficiência operacional entre os diferentes *stakeholders*.

O k-means é um algoritmo que treina um modelo para agrupar objetos semelhantes. Para isso, ele mapeia cada observação no conjunto de dados de entrada para um ponto no espaço de “n” dimensões (em que “n” é o número de atributos da observação). (AMAZON WEB SERVICES, [s.d.], n.p.).

2.4 MODELAGEM PREDITIVA E OTIMIZAÇÃO

O algoritmo XGBoost foi selecionado para prever o Lead Time. A otimização de *hiperparâmetros* foi realizada pelo Optuna, ajustando elementos como profundidade das árvores, número de estimadores e taxa de aprendizado.

Esse processo, conhecido como otimização de *hiperparâmetros*, tem o objetivo de melhorar o desempenho dos algoritmos por meio da seleção das combinações que produzem os resultados mais precisos.

Os *hiperparâmetros* são definições feitas antes do treinamento que controlam a arquitetura e o aprendizado do modelo. Eles não são aprendidos pelo algoritmo e influenciam diretamente sua capacidade de generalização.

Como ajustá-los é complexo, ferramentas como o Optuna automatizam a busca pelas melhores combinações para melhorar o desempenho do modelo (PINHEIRO, 2023).

Para garantir a transparência exigida em ambientes multissetoriais, utilizaram-se SHAP Values. Enquanto a importância das variáveis hierarquiza o peso de cada fator, os valores SHAP quantificam o impacto individual de variáveis burocráticas no *Lead Time* final. Esse recurso provê a prova científica necessária para que gestores identifiquem gargalos de anuência e atuem de forma proativa na governança do complexo.

3. RESULTADOS E DISCUSSÃO

A análise via PCA revelou que uma pequena parcela de variáveis críticas domina a variância operacional em Santos, reforçando que a integração via PCS simplifica a gestão.

O modelo XGBoost, após sintonização pelo Optuna, apresentou desempenho superior, demonstrando capacidade de prever atrasos com base no histórico de integração.

A clusterização identificou grupos com diferentes níveis de maturidade digital, validando a prática de "Capacitação e Engajamento".

A discussão dos resultados comprova que a fragmentação de dados é uma falha de governança corrigível com visibilidade analítica.

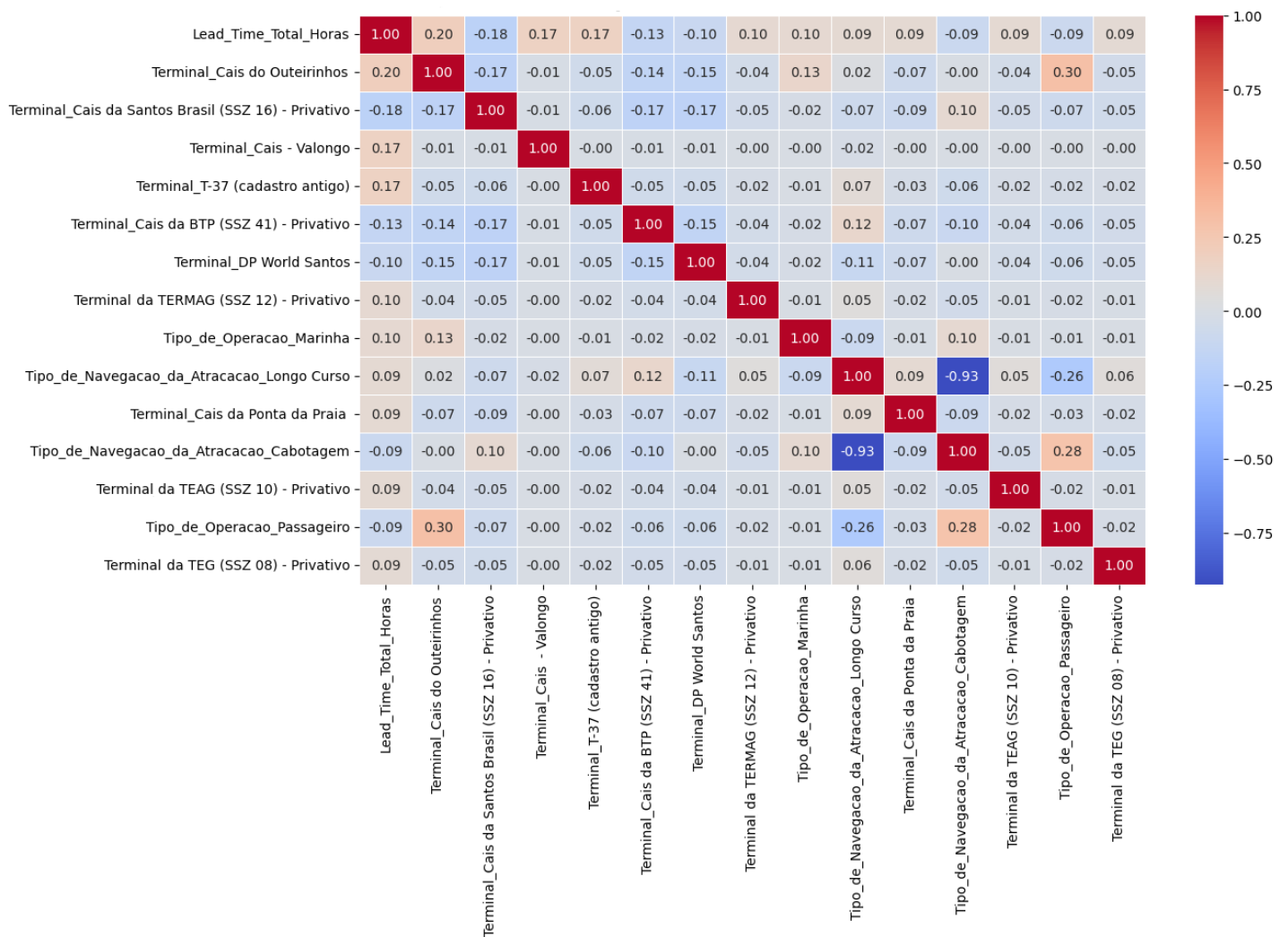
Esses agrupamentos evidenciam não apenas diferenças técnicas, mas posições distintas no campo portuário, nas quais o acesso ao capital informacional influencia diretamente a capacidade de integração ao PCS.

Para facilitar a interpretação dos resultados, foram gerados gráficos utilizando as bibliotecas Matplotlib (HUNTER, 2007) e Seaborn (WASKOM, 2021), ambas amplamente empregadas na comunidade Python para análise e visualização de dados.

A Matriz de Correlação de Pearson (Figura 2) forneceu o diagnóstico inicial das relações entre variáveis operacionais e o Lead Time.

A análise via PCA sintetizou a complexidade dos dados, explicando 50,73% da variância total nos dois primeiros eixos.

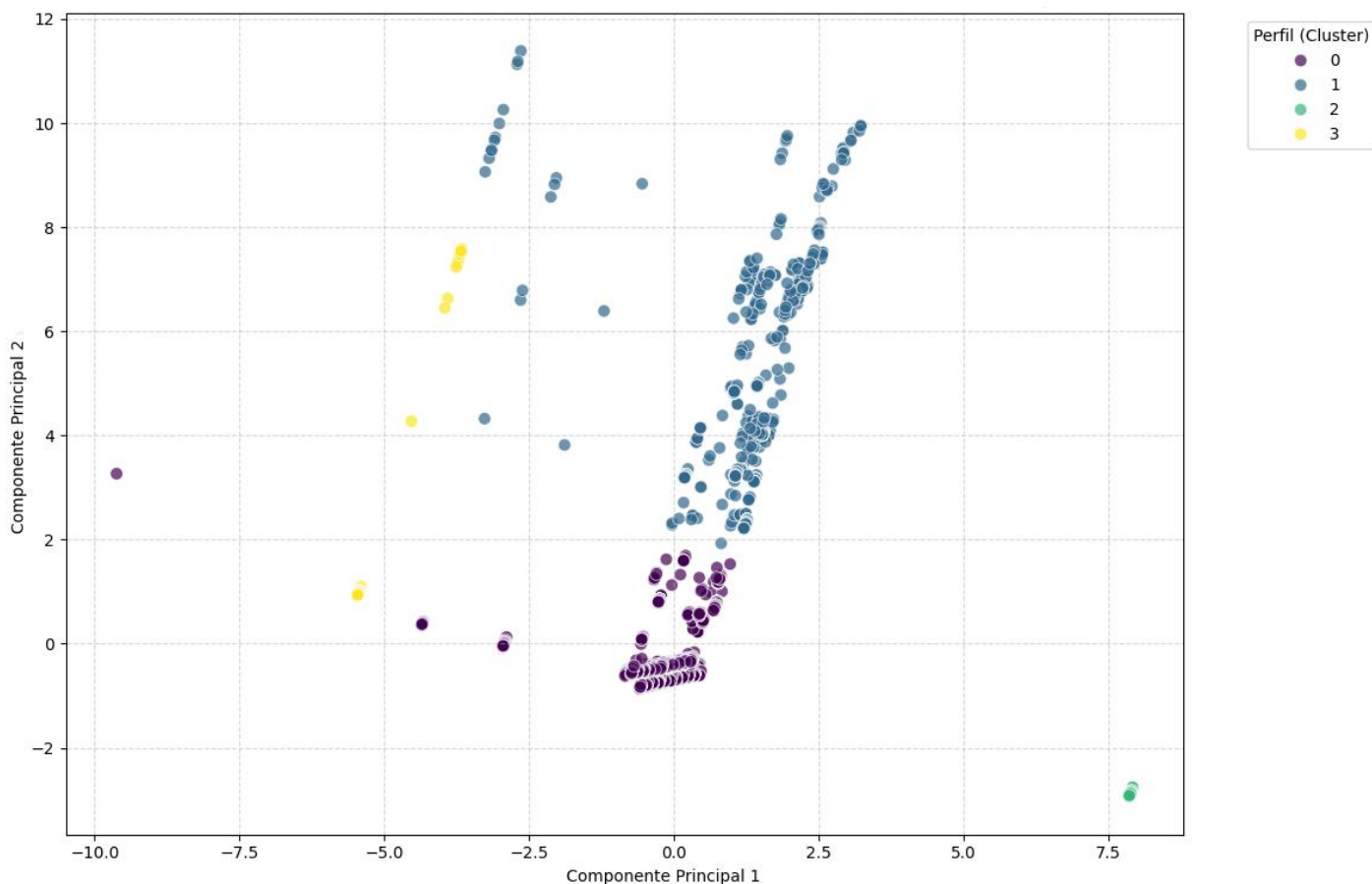
Figura 2 – Matriz de correlação de Pearson entre variáveis operacionais e o Lead Time.



Fonte: Feito no Google Colab pelos Autores (2025)

A Figura 3 segmentou as operações em quatro perfis de governança. O *cluster 2* mostrou-se o mais eficiente (112,9h), enquanto o *Cluster 1* apresentou o maior gargalo (147h). Isso valida que o acesso ao capital informacional influencia diretamente a eficiência no complexo santista.

Figura 3 – Agrupamento das operações portuárias

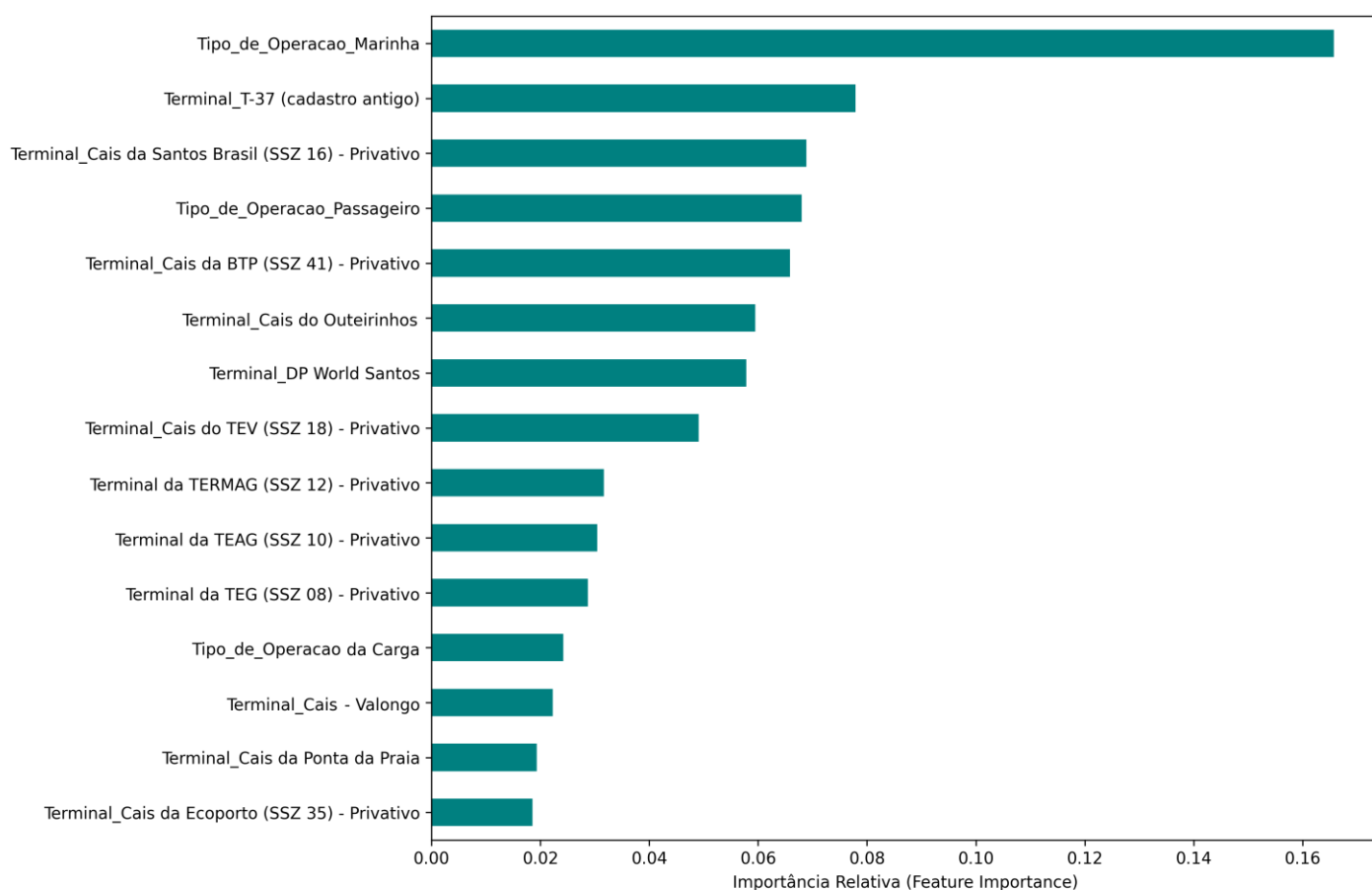


Fonte: Feito no Google Colab pelos Autores (2025)

A Figura 4 apresenta a hierarquia dos fatores determinantes segundo o modelo XGBoost. Esta análise revela quais variáveis operacionais e burocráticas possuem maior peso estatístico na composição do Lead Time no Porto de Santos.

A predominância de indicadores relacionados aos tempos de processamento externo confirma que a eficiência não depende apenas da infraestrutura física, mas da fluidez do fluxo informacional entre os diversos atores da comunidade portuária.

Figura 4 – Ranking de importância das variáveis (Feature Importance) do modelo XGBoost.

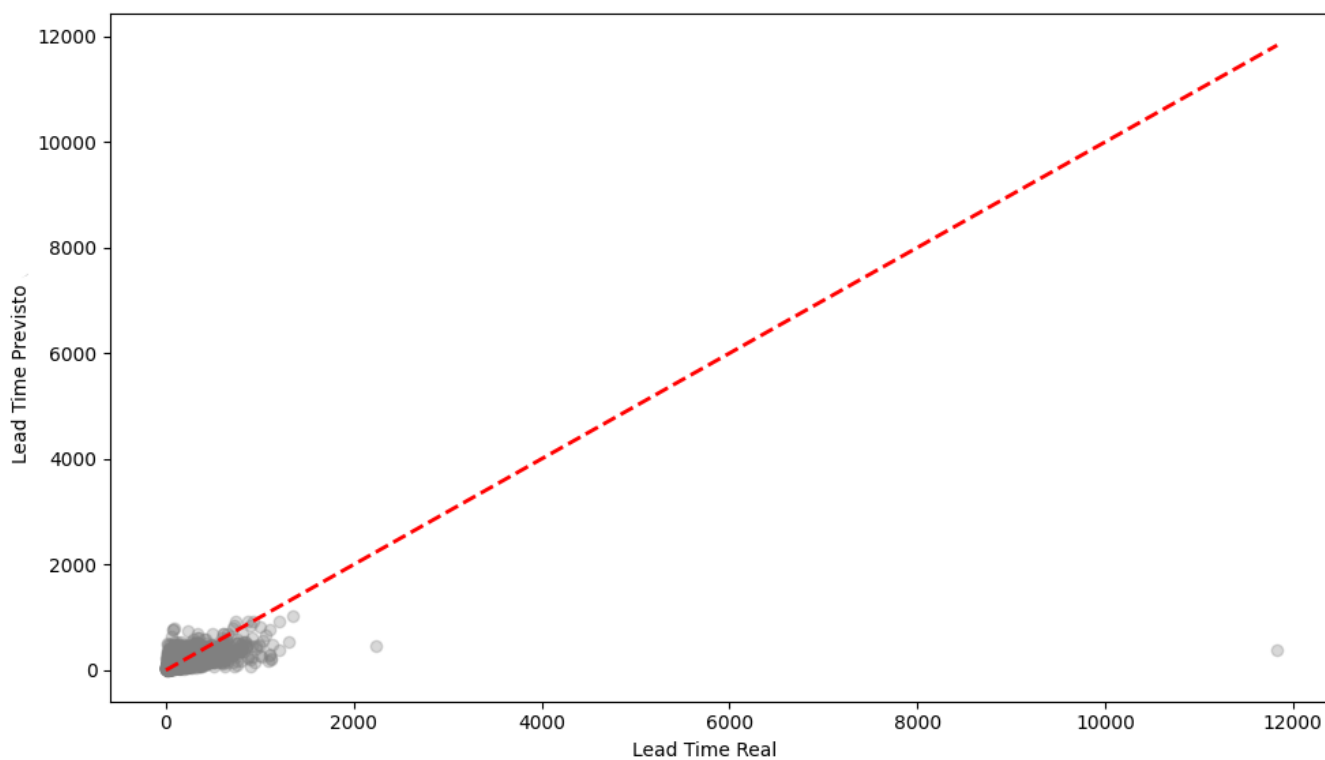


Fonte: Feito no Google Colab pelos Autores (2025)

O modelo apresentou um coeficiente de determinação (R^2) de 0,22 e um Erro Médio Absoluto (MAE) de 78,98 horas. Tais métricas, embora demonstrem uma capacidade preditiva moderada, confirmam estatisticamente a elevada entropia e fragmentação informacional do Porto de Santos.

O baixo R^2 valida a hipótese de que variáveis externas não digitalizadas e a falta de padronização entre anuentes inserem ruídos críticos na cadeia, reforçando que a viabilidade de um PCS é, antes de tudo, uma necessidade de saneamento de dados para reduzir a volatilidade operacional.

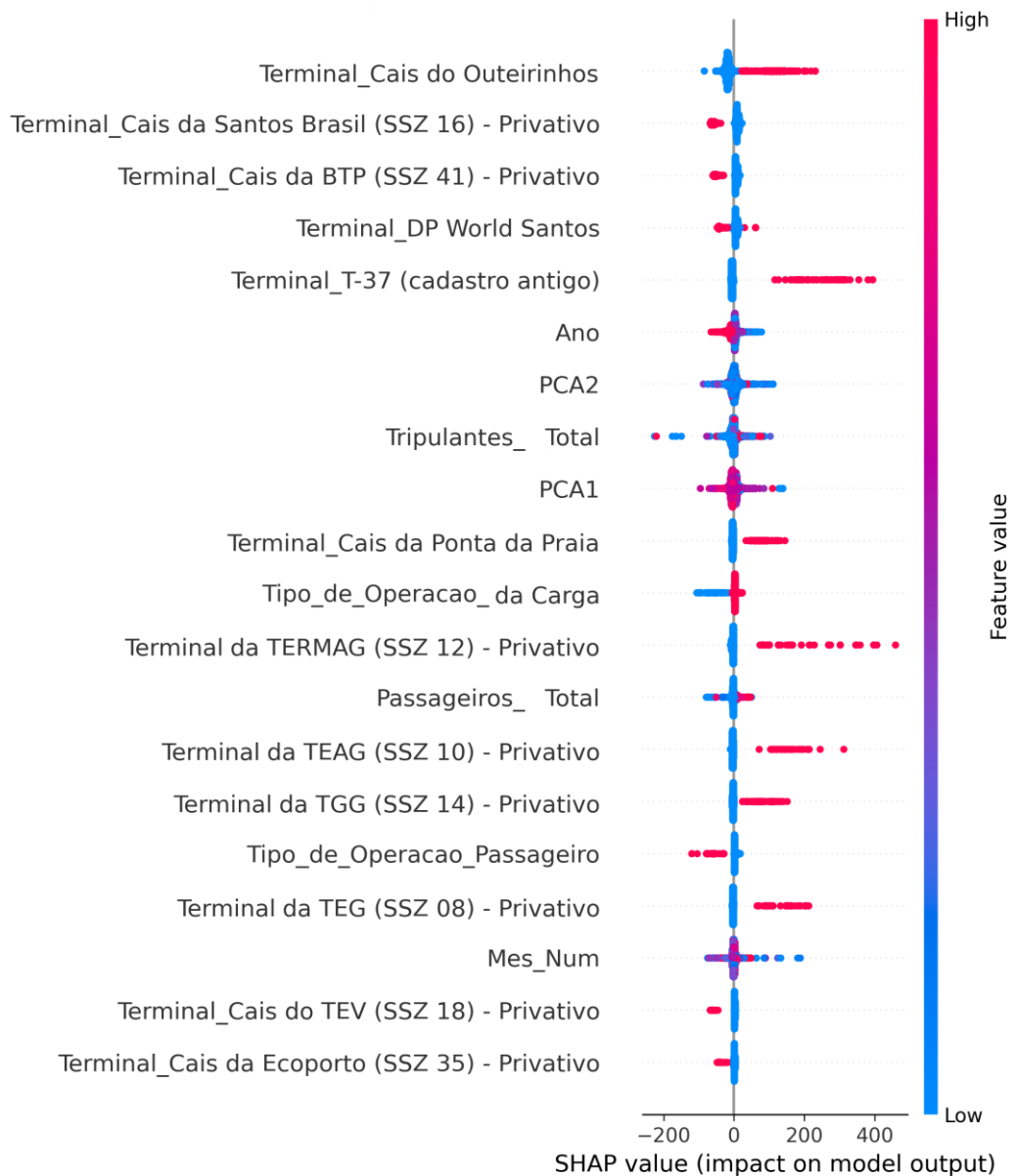
Figura 5 – Dispersão entre Lead Time real vs. predito e coeficiente de determinação (R^2).



Fonte: Feito no Google Colab pelos Autores (2025)

A Figura 6, baseada nos SHAP Values, aprofunda a compreensão sobre a direção do impacto de cada variável. A análise de Shapley confirmou que valores elevados de burocracia (especialmente os tempos médios de análise de órgãos anuentes) impactam positivamente o incremento do Lead Time, deslocando a operação para uma zona de ineficiência.

Figura 6 – Distribuição dos valores SHAP para explicabilidade do impacto das variáveis



Fonte: Feito no Google Colab pelos Autores (2025)

Este resultado fornece a prova científica de que a fragmentação de dados entre órgãos reguladores é um gargalo de governança que um PCS pode mitigar ao oferecer visibilidade analítica em tempo real e integração de processos.

4. CONSIDERAÇÕES FINAIS

Este artigo conclui a trilogia de estudos sobre a modernização do Porto de Santos ao aplicar técnicas avançadas de Ciência de Dados ao cenário real do complexo santista. A utilização dos modelos XGBoost e Optuna provou ser essencial para conferir capacidade analítica avançada à governança digital, transformando dados brutos integrados da ANTAQ e Porto Sem Papel em argumentos técnicos para a colaboração multissetorial.

Os resultados demonstram que a viabilidade de um PCS em Santos transcende a infraestrutura tecnológica, dependendo da união entre uma estrutura organizacional robusta e a capacidade preditiva. O modelo identificou que a fragmentação informacional entre órgãos reguladores e entes privados é o principal gargalo de governança, mas que a transparência analítica (via SHAP Values) oferece a prova científica necessária para mediar conflitos de interesse e reduzir a resistência cultural ao compartilhamento de dados.

Como contribuição prática, este estudo propõe um *framework* onde a análise de dados fundamenta a tomada de decisão compartilhada, elevando a competitividade do complexo.

Como próximo passo, recomenda-se o desenvolvimento de um motor de inferência que integre os resultados preditivos a Dashboards de Business Intelligence (BI). Tal implementação permitiria uma Governança Preditiva em tempo real, na qual os gestores poderiam identificar proativamente gargalos burocráticos e atuar na redução dos tempos de anuência antes que estes impactem o Lead Time final do Porto de Santos.

REFERÊNCIAS

ANTAQ – AGÊNCIA NACIONAL DE TRANSPORTES AQUAVIÁRIOS. **Anuário Estatístico Aquaviário**. Brasília, 2024. Disponível em: <https://www.gov.br/antaaq>. Acesso em: 15 fev. 2026.

AMAZON WEB SERVICES. **Como funciona o clustering do k-means**. [S.l.]: Amazon Web Services, [s.d.]. Disponível em: https://docs.aws.amazon.com/pt_br/sagemaker/latest/dg/algo-kmeans-tech-notes.html. Acesso em: 13 nov. 2025.

BRASIL. Ministério de Portos e Aeroportos. **Porto Sem Papel (PSP)**: Documento Único Virtual (DUV). Brasília, 2024. Disponível em: <https://www.gov.br/portos-e-aeroportos/pt-br/assuntos/porto-sem-papel>. Acesso em: 15 fev. 2026.

CHARLEAUX, Lupa; TOLEDO, Victor. **O que é Machine Learning?** Tecnoblog, out. 2024. Disponível em: <https://tecnoblog.net/responde/machine-learning-o-que-e-como-funciona-e-quais-sao-os-tipos-de-aprendizado-de-maquina/>. Acesso em: 10 nov. 2025.

HASTIE, Trevor; TIBSHIRANI, Robert; FRIEDMAN, Jerome. **The elements of statistical learning: data mining, inference, and prediction**. 2. ed. New York: Springer, 2009.

HUNTER, John D. Matplotlib: A 2D graphics environment. **Computing in Science & Engineering**, v. 9, n. 3, p. 90-95, 2007.

I JUNIOR. **Automação de processos: como reduzir o retrabalho gerado por tarefas manuais**. [S.l.], 22 set. 2022. Disponível em: <https://ijunior.com.br/automacao-de-processos-como-reduzir-o-retrabalho-gerado-por-tarefas-manuais/>. Acesso em: 27 fev. 2026.

IBM. **What is Principal Component Analysis (PCA)?** [S. l.], 2025. Disponível em: <https://www.ibm.com/topics/principal-component-analysis>. Acesso em: 11 nov. 2025.

MINISTÉRIO DE PORTOS E AEROPORTOS. **Porto de Santos está entre organizações mais eficientes do país, segundo publicação nacional**. Brasília: MPor, 09 dez. 2025. Disponível em: <https://www.gov.br/portos-e-aeroportos/pt-br/assuntos/noticias/2025/12/porto-de-santos-esta-entre-organizacoes-mais-eficientes-do-pais-segundo-publicacao-nacional>. Acesso em: 27 fev. 2026.

PEDREGOSA, F. et al. Scikit-learn: Machine Learning in Python. **Journal of Machine Learning Research**, v. 12, p. 2825-2830, 2011.

PINHEIRO, João Manoel Herrera. **Um estudo sobre Algoritmos de Boosting e a Otimização de Hiperparâmetros Utilizando Optuna**. 2023. 147 p. Monografia (Engenharia Mecatrônica) – Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2023.

SIMÕES, Priscila Bayer de Oliveira et al. **Benefícios da implantação de Port Community Systems: uma revisão da literatura**. In: CONGRESSO NACIONAL INTEGRA PORTOS (CNIT), 2024.

SIMÕES, Priscila Bayer de Oliveira et al. **Identificação de boas práticas em Port Community Systems: uma revisão de literatura**. In: XLV ENCONTRO NACIONAL DE ENGENHARIA DE PRODUÇÃO (ENEGEP), Natal, 2025.

WASKOM, Michael L. Seaborn: statistical data visualization. **Journal of Open-Source Software**, v. 6, n. 60, p. 3021, 2021.

YOKOYAMA, Naoki. **Modelos de Machine Learning**. Medium, 30 out. 2020. Disponível em: <https://naokiyokoyama.medium.com/modelos-de-machine-learning-bcb3f8ed1513>. Acesso em: 15 nov. 2025